CLAIMS

What is claimed is:

1. A method for processing a signal, the method comprising the steps of:
dividing the signal into frames, each frame having a corresponding spectrum;
selecting a plurality of pitch candidates from a first frame;
selecting a plurality of pitch candidates from a second frame;
calculating a cumulative error function for a plurality of paths, each path including a pitch candidate from the first frame and a pitch candidate from the second frame;
selecting a path corresponding to a low cumulative error function;
basing a pitch estimate for a current frame on the selected path;
using the pitch estimate for the current frame to process the signal.

2. The method of claim 1, wherein the determining step further comprises the step of synthesizing the signal, the synthesized signal having corresponding spectra, wherein the pitch candidates are obtained from the synthesized spectra.

3. The method of claim 1 wherein the first frame is a previous frame and the second frame is a current frame.

4. The method of claim 1 wherein the first frame is a current frame and the second frame is a future frame.

5. The method of claim 1 wherein the plurality of pitch candidates for the first frame is no more than five pitch candidates and the plurality of pitch candidates for the second frame is no more than five pitch candidates.

6. The method of claim 5 wherein a cumulative error function is calculated for all possible paths.

7. The method of claim 1 wherein the selected pitch candidates for the first and second frames have low error functions.

8. The method of claim 7 wherein the error function is a measure of the spectral error between original and synthesized spectra.

9. The method of claim 1 further comprising the step of selecting a plurality of pitch candidates for a third frame, and wherein each path further includes a pitch candidate from the third frame.

10. The method of claim 9 wherein the plurality of pitch candidates for the first frame is no more than five pitch candidates, the plurality of pitch candidates for the second frame is no more than five pitch candidates and the plurality of pitch candidates for the third frame is no more than five pitch candidates.

11. The method of claim 10 wherein a cumulative error function is calculated for all possible paths.

12. The method of claim 9 wherein the first frame is a previous frame, the second frame is a current frame and the third frame is a future frame.

13. The method of claim 9 wherein the selected pitch candidates for the first, second and third frames have low error functions.

14. The method of claim 13 wherein the error function is a measure of the spectral error between original and synthesized spectra.

15. The method of claim 14 wherein a cumulative error function for each path is defined by the equation:

$$CF = k * (E_{-1} + E_{-2}) + \log (P_{-1} / P_{-2}) + k * (E_{-2} + E_{-3}) + \log(P_{-2} / P_{-3})$$

wherein P_{-1} is a selected pitch candidate for the first frame, P_{-2} is a selected pitch candidate for the second frame, P_{-3} is a selected pitch estimate for the third frame, E_{-1} is an error for P_{-1}, E_{-2} is an error for P_{-2}, E_{-3} is an error for P_{-3}, and k is a penalising factor.

16. The method of claim 9 wherein the basing a pitch estimate for a current frame on the selected path step further comprises calculating a backward pitch estimate along the selected path, wherein the pitch estimate for a current frame is based on the selected path and the backward pitch estimate.

17. The method of claim 16 wherein the backward pitch estimate is calculated by calculating backward sub-multiples of a pitch candidate for the second frame in the selected path, determining whether the backward sub-multiples satisfy backward constraint equations, and selecting a low backward sub-multiple as the backward pitch estimate wherein the pitch candidate for the second frame in the selected path is selected as the backward pitch estimate if a backward sub-multiple does not satisfy the backward constraint equations.

18. The method of claim 17 wherein the basing a pitch estimate for a current frame on the selected path step further includes determining a backward cumulative error based on the backward pitch estimate.

19. The method of claim 18, wherein the backward cumulative error is defined by:

$$CE_B(P_B) = E(P_B) + E_{-1}(P_{-1})$$

wherein $E(P_B)$ is an error of the backward pitch estimate and $E_{-1}(P_{-1})$ is an error of the first pitch candidate.

20. The method of claim 9 wherein the basing a pitch estimate for a current frame on the selected path step further comprises calculating a forward pitch estimate along the selected path, wherein the pitch estimate for a current frame is based on the selected path and the forward pitch estimate.

21. The method of claim 20 wherein the forward pitch estimate is calculated by calculating forward sub-multiples of a pitch candidate for the second frame in the selected path, determining whether the forward sub-multiples satisfy forward constraint equations, and selecting a low forward sub-multiple as the forward pitch estimate wherein the pitch

candidate for the second frame in the selected path is selected as the forward pitch estimate if a forward sub-multiple does not satisfy the forward constraint equations.

22. The method of claim 21 wherein the forward constraint equation is selected from the group consisting of:

$CE_F(P_0 / n) \leq 0.85$ and $(CE_F(P_0 / n)) / (CE_F(P_0)) \leq 1.7$;

$CE_F(P_0 / n) \leq 0.4$ and $(CE_F(P_0 / n)) / (CE_F(P_0)) \leq 3.5$; and

$CE_F(P_0 / n) \leq 0.5$

where P_0 / n refers to forward sub-multiples, P_0 refers to the pitch candidate for the second frame in the selected path, and $CE_F(P)$ is an error function.

23. The method of claim 21 wherein the basing a pitch estimate for a current frame on the selected path step further includes determining a forward cumulative error based on the forward pitch estimate.

24. The method of claim 23, wherein the forward cumulative error is defined by:

$CE_F(P_F) = E(P_F) + E_{-1}(P_{-1})$

wherein $E(P_F)$ is an error for the forward pitch estimate and $E_{-1}(P_{-1})$ is an error of the first pitch candidate.

25. The method of claim 24 wherein the basing a pitch estimate for a current frame on the selected path step further comprises calculating a backward pitch estimate along the selected path, wherein the backward pitch estimate is used to calculate a backward cumulative error, the pitch estimate being based on the selected path, the forward cumulative error and the backward cumulative error.

26. The method of claim 25, wherein the basing a pitch estimate for a current frame on the selected path step further comprises comparing the forward and backward cumulative errors with one another, selecting the pitch estimate as the forward pitch estimate if the forward cumulative error is less than the backward cumulative error, and selecting the

pitch estimate as the backward pitch estimate if the backward cumulative error is less than the forward cumulative error.

27. The method of claim 20 wherein the basing a pitch estimate for a current frame on the selected path step further comprises calculating a backward pitch estimate along the selected path, wherein the pitch estimate for a current frame is based on the selected path, the forward pitch estimate and the backward pitch estimate.

28. A method for processing a signal comprising the steps of:

dividing the signal into frames;

obtaining a pitch estimate for a current frame;

refining the obtained pitch estimate comprising the sub-step of:

computing backward and forward sub-multiples of the obtained pitch estimate for the current frame;

determining whether the backward sub-multiples satisfy at least one backward constraint equation;

determining whether the forward sub-multiples satisfy at least one forward constraint equation;

selecting a low backward sub-multiple that satisfies the at least one backward constraint equation as the backward pitch estimate, wherein the obtained pitch estimate of the current frame is selected as the backward pitch estimate if a backward sub-multiple does not satisfy the at least one backward constraint equation;

selecting a low forward sub-multiple that satisfies the at least one forward constraint equation as the forward pitch estimate, wherein the obtained pitch estimate of the current frame is selected as the forward pitch estimate if a forward sub-multiple does not satisfy the at least one forward constraint equation;

using the backward pitch estimate to compute a backward cumulative error;

using the forward pitch estimate to compute a forward cumulative error;

comparing the forward cumulative error to the backward cumulative error;

refining the chosen pitch estimate for the current frame based on the comparison; and

using the refined pitch estimate for the current frame to process the signal.

29. A method for making voicing decisions for segments of a signal to process the signal, comprising the steps of:

dividing the signal into frames, each frame having a corresponding spectrum;

tracking a base noise energy level of previous frames;

computing energy in a frame;

calculating a ratio between the energy of the frame and the base noise energy level;

comparing the ratio against a threshold value; and

declaring the frame unvoiced if the ratio is less than the threshold value, wherein the frame is declared voiced if the ratio is greater than the threshold value; and

using the declaration to process the signal.

30. The method of claim 29, wherein the threshold value is derived from heuristics.

31. The method of claim 30, wherein the heuristics are obtained from testing a set of about 10,000 to about 15,000 frames with different background noise levels.

32. The method of claim 29 further comprising the steps of:

dividing a spectrum of a frame previously declared voiced into bands;

comparing at least one band of the spectrum with at least one band of a voiced synthesized spectrum;

comparing at least one band of the spectrum with at least one band of an unvoiced synthesized spectrum;

making a voiced/unvoiced decision for the at least one band based on the original-voiced comparison and the original-unvoiced comparison; and

redeclaring the frame unvoiced if each band of the frame is marked unvoiced.

33. The method of claim 32, wherein each band contains about three harmonics.

34. The method of claim 32, wherein the dividing the signal into frames step is based on a pitch frequency of the frame.

35. The method of claim 32, further comprising the steps of:

computing an unvoiced frame's energy;

comparing the unvoiced frame's energy with an empirical threshold value; and

redeclaring the unvoiced frame as silent if the frame has an energy level below the empirical threshold value.

36. The method of claim 35, wherein a voicing parameter is determined, the voicing parameter being used to transmit or store the voiced/unvoiced/silence band information.

37. A method to process a signal comprising the steps of:

using a first technique to synthesize an original spectrum of the signal;

using a second technique to synthesize the original spectrum;

calculating a first error between a band of the original spectrum and a corresponding band of the first synthesized spectrum;

calculating a second error between the band of the original spectrum and a corresponding band of the second synthesized spectrum;

comparing the first error to the second error;

declaring the band of the original spectrum as a first category band if the first error is less than the second error;

declaring the band of the original spectrum as a second category band if the second error is less than the first error; and

using the declaration to process the signal.

38. The method of claim 37, wherein the first technique comprises the step of ensuring that a valley amplitude between two successive harmonics of the first synthesized spectrum is not less than a valley amplitude between two corresponding successive harmonics of the original spectrum.

39. The method of claim 38, wherein the first technique further comprises the step of clipping the synthesized spectrum valley amplitude to about a minimum value of the corresponding valley amplitude of the original spectrum.

40. The method of claim 39, wherein the first technique further comprises the steps of placing a window around harmonic amplitudes to determine a pitch frequency.

41. The method of claim 37, wherein the second technique comprises the step of fixing a pitch frequency and calculating a root mean square value over a region in the spectrum.

42. The method of claim 37, wherein the error calculated between the band of the original spectrum and the corresponding band of the first synthesized spectrum is a mean square error, and the error calculated between the band of the original spectrum and the corresponding band of the second synthesized spectrum is a mean square error.

43. The method of claim 37, wherein the original spectrum of the signal corresponds to a frame marked voiced.

44. The method of claim 43, wherein the first synthesizing technique is performed assuming a band of the spectrum is voiced, the second synthesizing technique is performed assuming a band of the spectrum is unvoiced, the band is declared voiced if the first error is

less than the second error, and the band is declared unvoiced if the second error is less than the first error, wherein the using a first and second technique to synthesize steps, the calculating steps, the comparing step, the declaring step, the using the declaration to process step are performed for each band of the spectrum, and a frame corresponding to the spectrum is declared unvoiced if each band within the frame is declared unvoiced.

45. The method of claim 37, wherein the error of the first technique is defined by the equation:

$$error_{1st}(k) = [(S_{org}(m) - S_{synth}(m)) * (S_{org}(m) - S_{synth}(m))] / N$$

where k is the band number, $S_{org}(m)$ is the original spectrum, $S_{synth}(m)$ is the first synthesized spectrum, and N is a number of points used over a region to calculate a mean square error.

46. The method of claim 37, wherein the error of the second technique is defined by the equation:

$$error_{2nd}(k) = [(S_{org}(m) - S_{rms}(m)) * (S_{org}(m) - S_{rms}(m))] / N$$

where k is the band number, $S_{org}(m)$ is the original spectrum, $S_{rms}(m)$ is the second synthesized spectrum, and N is a number of points used over a region to calculate a mean square error.

47. The method of claim 37, wherein the first synthesizing technique is performed assuming each band of the spectrum is voiced, the second synthesizing technique is performed assuming each band of the spectrum is unvoiced, the band is declared voiced and the voiced synthesis form is used if the first error is less than the second error, and the band is declared unvoiced and the unvoiced synthesis form is used if the second error is less than the first error; and further comprising the steps of:

determining a voicing parameter, wherein the voicing parameter denotes a band threshold and is transmitted or stored to convey voiced/unvoiced band information.

48. The method of claim 47, wherein bands having properties below the voicing parameter are declared unvoiced and bands having properties above the voicing parameter are declared voiced.

49. The method of claim 47, wherein the voicing parameter is determined to minimize a hamming distance between a voicing bit string of the original band and a voicing bit string of the synthesized band.

50. The method of claim 47, wherein the voicing parameter is weighted to compensate for voiced bands that are declared unvoiced and for unvoiced bands that are declared voiced by previous unvoiced/voiced band declarations.

51. The method of claim 47, wherein the following weighted bit error function is applied to the voicing parameter:

$$\varepsilon(k) = c_v \sum_{i=1}^{k} (1-a_i) + \sum_{j=k+1}^{m} a_j$$

wherein a_i, $i = 1, \ldots, m$ are previous band declarations, c_v is a constant, and k is the harmonic number of the spectrum.

52. A method of transmitting voicing decisions for bands of signal frames, comprising the steps of:

determining a voicing parameter for which a distance between an original band voicing bit stream and a synthesized band voicing bit stream is minimized;

declaring all bands having properties below the voicing parameter as unvoiced and all bands having properties above the voicing parameter as voiced; and,

transmitting the voicing parameter.

53. The method of claim 52 further comprising the steps of quantizing and encoding the voicing parameter.

54. The method of claim 52, wherein the following weighted bit error function is applied to the voicing parameter:

$$\varepsilon(k) = c_v \sum_{i=1}^{k} (1-a_i) + \sum_{j=k+1}^{m} a_j$$

where a_i, $i = 1, \ldots, m$ are previous voicing decisions, c_v is a constant, and k is the harmonic number of the spectrum.

~~55.~~ A method of transmitting or storing signal information, comprising the steps of:

synthesizing a spectrum of a signal;

modeling spectral amplitudes of the spectrum using a linear prediction technique;

mapping linear prediction coefficients obtained from the linear prediction model to corresponding line spectral pairs;

quantizing the line spectral pairs, wherein a residual of a previous quantizing stage is quantized during a current quantizing stage;

storing or transmitting the quantized line spectral pairs.

56. The method of claim 55 wherein multi-stage vector quantization is used during the quantizing step.

57. The method of claim 55 further comprising the steps of:

determining a voicing parameter, the voicing parameter conveying voiced/unvoiced band information;

quantizing and encoding the voicing parameter; and

storing or transmitting the quantized voicing parameter.

~~58.~~ A method for synthesizing bands of signal frames previously declared unvoiced, comprising the steps of:

generating a random noise sequence;

transforming values of the random noise sequence to random phase values;

assigning the transformed random phase values to the spectral amplitudes to obtain a modified unvoiced spectrum; and

taking an inverse Fourier transform of the modified unvoiced spectrum to obtain an unvoiced speech signal.

59. The method of claim 58, wherein the random noise sequence is defined by values generated from the equation:

$$U(n+1) = 171 * U(n) + 11213 - 53125 * \lfloor (171 * U(n) + 11213) / 53125) \rfloor$$

wherein $\lfloor \ \rfloor$ represent the integer portion of a fractional number and $U(0)$ is initially set to 3147.

60. The method of claim 58, wherein the synthesizing step further includes the step of applying a weighted overlap add method to the unvoiced speech signal.

61. The method of claim 58 wherein the transformed values are assigned to random phase values between about negative π and about positive π.

62. The method of claim 58 wherein the transformed random phase values are assigned by applying the equation:

$$U_w(m) = S_{amp}(l) * (\cos(\phi) + j\sin(\phi))$$

where l is a harmonic of the unvoiced spectrum and ϕ is the random phase assigned to the lth harmonic of the unvoiced spectrum.

63. The method of claim 58 wherein the unvoiced speech signal is obtained by applying the inverse Fourier transform equation:

$$u(n) = 1/N * \sum_{m=-N/2}^{m=(N/2)-1} U(m)\exp((j*2*\pi*m*n)/N)$$

where $N/2 \leq n < (N/2) - 1$, and N is the number of points used in the computation.

64. A method for processing received data, comprising the steps of:

decoding the received data to obtain signal information including unvoiced frame information, voiced frame information, and spectral envelope information;

initializing phases of harmonics of the spectral envelope to a fixed set of values, wherein the initializing step is performed for transitions from unvoiced frames to voiced frames; and

65. The method of claim 64 wherein the initialized phases are related to get a balanced output speech waveform.

66. The method of claim 65 wherein the fixed set of values is approximately defined by

0.000000	-1.781785	-1.085355	-0.009230
-2.008388	-0.588607	-1.994504	-0.490472
-0.368968	-1.430010	-1.009797	-2.307083
-0.967567	0.650081	-1.705878	-0.299851
-2.077636	-0.414668	-0.127823	-1.878630
-0.699374	-1.733431	-2.051089	-0.088243
-0.853920	0.060934	-1.063803	-1.295832
-0.129658	-0.631376	-0.903947	-0.903947
0.425315	-2.315562	-0.903947	-0.903947
-0.897955	-0.901923	-0.903947	-0.903947
0.511909	-2.362951	-0.903947	-0.903947
-0.957640	-0.926879	-0.903947	-0.903947
-2.185920	0.495461	-0.903947	

wherein about 0.000000 is the first fixed value used and about -2.008388 is the second fixed value used so that each succeeding value is the next value used.

67. An encoder for processing a signal comprising:

a means for making a voicing decision for a frame of the signal;

a means for determining a pitch value for frames marked voiced;

a means for basing an unvoiced-voiced decision for bands of frames marked voiced on two error functions, the first error function comprising a difference between a voiced synthesized spectrum and a spectrum of the signal and the second error function comprising a difference between an unvoiced synthesized spectrum and the spectrum of the signal;

a means for synthesizing frames marked unvoiced;

a means for synthesizing frames marked voiced;

a means for quantizing signal information.

68. The encoder of claim 67 further comprising a means for storing signal information.

69. The encoder of claim 67 further comprising a means for transmitting signal information.

70. The encoder of claim 67 wherein the means for determining a pitch value comprises: (a) a means for selecting pitch candidates from a first frame and a second frame, (b) a means for calculating an error for a plurality of paths, each path including a pitch candidate from the first frame and a pitch candidate from the second frame, and (c) a means for basing a pitch value on a path having a low error.

71. The encoder of claim 67 further comprising a means for refining the pitch value.

72. The encoder of claim 67 wherein frames marked voiced are divided into bands based on a pitch value.

73. The encoder of claim 67 further comprising a means for modeling spectral amplitudes of synthesized spectra, the means for modeling synthesized spectra using a linear prediction technique, and a means for converting linear prediction coefficients to line spectral pairs.

74. The encoder of claim 67 further comprising a means for calculating a voicing parameter, the voicing parameter being weighted to compensate for voiced bands marked unvoiced and for unvoiced bands marked voiced.

75. The encoder of claim 74 wherein the means for determining a pitch value comprises: (a) a means for selecting pitch candidates from a first frame and a second frame,

(b) a means for calculating an error for a plurality of paths, each path including a pitch candidate from the first frame and a pitch candidate from the second frame, and (c) a means for basing a pitch value on a path having a low error;

the means for basing an unvoiced-voiced decision for bands of frames marked voiced on two error function comprising a means for dividing frames marked voiced into bands based on the pitch value; and the encoder further comprises:

a means for refining the pitch value;

a means for modeling spectral amplitudes of synthesized spectra, the means for modeling synthesized spectra using a linear prediction technique; and

a means for converting linear prediction coefficients to line spectral pairs.

76. The encoder of claim 67 wherein the means for making a voicing decision, the means for determining a pitch value, the means for basing a unvoiced-voiced decision for bands of frames marked voiced on two error functions, the means for synthesizing frames marked voiced, the means for synthesizing frames marked unvoiced, and the means for quantising signal information is a speech coder or a speech vocoder.

77. A decoder for processing received data, comprising:

a means for decoding received data;

a means for synthesizing frames having unvoiced band information to produce unvoiced speech, the means for synthesizing frames having unvoiced band information including a random number generator, means for transforming random values from the random number generator to random phase values, means for assigning the random phase values to spectral amplitudes of a spectral shape vector to form a modified unvoiced spectrum, and means for taking an inverse Fourier transform of the modified unvoiced spectrum;

a means for synthesizing frames having voiced band information to produce voiced speech; and

a means for combing voiced speech and unvoiced speech.

78. The decoder of claim 77, wherein unvoiced band information, voiced band information, line spectral pair information, and pitch value information is decoded, the decoder further including means for converting line spectral pair information to a spectral shape vector.

79. The decoder of claim 77, further including a means for initializing phases of harmonics of unvoiced bands to a fixed set of values, wherein the initializations are performed for transitions from unvoiced frames to voiced frames.